An Analysis of What Should Be Easily Preventable Data Breaches

Chlotia P. Garrison, Winthrop University,
316 Thurmond Bldg, Rock Hill SC 29733

Moeti Ncube, Nextera Energy Resources,
700 Universe Blvd EPM/JB, Juno Beach, FL 33404

Abstract

While defending against hacking is important, businesses should not ignore what should be easily preventable security incidents. Studies reveal that accidents cause one-fourth to one-third of breaches of personal information. This study analyzes breach incidents that occurred through unintended disclosure and discarded non-electronic data. The results reveal there is a significant difference in the number of incidents per institution and per subtype. Education and Government had the highest number of incidents and most breaches are exposed via the Internet.

Introduction

In April 2011, the public became aware of two breaches of personal identifying information (PII) that affected millions of individuals. Sony, PlayStation Network, revealed that hackers had accessed their systems and obtained names, addresses, email addresses, dates of birth, passwords, billing address, and password security questions. Subsequent reports revealed that 12,000,000 unencrypted credit card numbers were easily accessible by the hackers. The Texas Comptroller's Office discovered that names, social security numbers, addresses, dates of birth and driver's license numbers, of 3.5 million Texans were available unencrypted on a public server for approximately a year (Privacy Rights Clearinghouse 2011; ITRC 2011). In both instances, the data was unencrypted. Unencrypted sensitive data on a private network could be considered unwise, whereas, unencrypted sensitive data on a public server might be considered irresponsible. Class action lawsuits have been filed in both cases (Privacy Rights Clearinghouse 2011). Preventing a hacking breach and preventing a breach from data placed on a public server are in two vastly different categories. Preventing a hacking breach requires technical expertise. It requires ever increasing technical expertise as hackers are continually circumventing the techniques of security

professionals or devising new attack methods. Sony and Epsilon, an e-mail marketing company, both experienced large breaches in 2011. According to Talbot, both companies indicated they were meeting industry security standards but were still victims of (multiple in the case of Sony) hacking breaches (Talbot 2011).

Preventing a data breach such as the one that occurred at the Texas Comptroller's Office requires little or no technical expertise. One would expect that non-technical prevention methods would be widely used and that breaches of this type would be rare. Unfortunately, this is not necessarily true. A study of data breaches between 2005 and 2009 revealed that 29% of data breaches at universities were of "unprotected data that may be publicly accessible" and includes breaches through e-mail, mail, internet and disposal (Ncube and Garrison 2010). Open Security Foundation's DataLossDB (2011-1), which has tracked breaches since 2001, has classified 23 % of all data breaches through 2010 as inside accidental, and an additional 3% as mishandling by an Insider. These percentages do not include Insider Malicious.

The public release of personal identifying information can place an individual at risk of identity theft or fraud. In 2010, the Consumer Sentinel database received 949,000 fraud & ID theft complaints, 250,854 were identify theft (FTC 2011-1). The Consumer Sentinel is an online database of consumer complaints available only to law enforcement (FTC nd). According to Javelin Strategy and Research (2011), 8.1 million Americans were victims of identity fraud in 2010.

The damage that can be caused by identity theft is highlighted by the emphasis placed on it by the Federal Trade Commission (FTC). The FTC (2011-2) identified, in its 2012 Budget Justification, a three-fold measure it uses to combat identity theft: bring law enforcement actions against companies that do not maintain reasonable safeguards to protect consumer information from identity theft; educate local law enforcement on identity theft; and educate consumers on how to avoid and recover from identity theft.

To help educate consumers, multiple organizations track the occurrence of breaches of personal identifying information (PII) also known as personal identification information and personally identifiable information. The Identity Theft Resource Center (ITRC) provides resources for victims of identity theft and maintains a list of breaches of PII that could lead to identity theft. The ITRC identifies five data loss methods: data on the move, accidental exposure, insider theft, subcontractors, hacking (ITRC 2011). The Open Security Foundation's DataLossDB (2011-2) identifies world-wide

data losses and provides statistics in 20 breach types including multiple categories of disposal, multiple categories of stolen, multiple categories of loss, multiple categories of missing, hack, e-mail, snail main, virus, fraud social engineering, and web. The Privacy Rights Clearinghouse (PRC) (2011) tracks breaches reported in the United States and uses the following breach types: Unintended disclosure, Hacking or malware, Insider, Physical loss, Portable device, Stationary device, and Unknown or other. Unintended disclosure encompasses information publicly posted on a website, mishandled or sent to the wrong party via email, fax or mail. Physical loss includes Lost, discarded or stolen non-electronic records.

The Study

This study analyzes the Privacy Rights Clearinghouse (PRC) data for breaches that should not occur. These breaches compromise PII through accidents, negligence, and mishandling. The PRC includes a brief description of each breach in its listing; the ITRC does not. Two categories of the PRC data were analyzed and used in this study, Unintended disclosure and Physical loss (Privacy Rights Clearinghouse 2011). This study includes all breaches identified as Unintended disclosure and those caused by discarded non-electronic records in Physical loss. The study excludes breaches of lost or stolen non-electronic records in the PRC Physical loss category. This study reviews a six-year period of breaches, 2005-2010, as recorded by the PRC. The data is analyzed by type of institution, state, and year publically reported. We consider the questions: What institutions are more likely to have an exposure? Is the number of exposures decreasing? Are businesses that have these breaches more likely to be located or have headquarters in certain states? Is a particular subcategory of exposure more likely to occur or more likely to occur to a particular institution type? Have some organizations had multiple incidents?

The Data

The Privacy Rights Clearinghouse data contained 591 exposure incidents. Of these, 229 had unknown, indeterminate, or zero records exposed. Incidents with zero records indicate a breach of data occurred but no social security or financial data was exposed. In some incidents, the breach notice provided insufficient information to determine if sensitive information was breached. Some breaches compromised PII, but it could not be determined how many records were breached. The PRC does not include these records in its total number of records breached. This analysis is based on the 362 incidents with a specific number of breached records identified. These 362 incidents

breached 13,010,340 records of PII. The number of breached records is not equal the number of individuals that had their information compromised. An individual could have multiple records containing their information exposed by the same breach. For example, the 2011 Epsilon breach affected several companies including Target, Best Buy, JPMorgan Chase and Marriott (Morran 2011; Lennon 2011); many were customers of multiple of these businesses. The data is recorded by the year the breach was disclosed to the public. The breach itself could have occurred in an earlier year. Some companies have reported breach information immediately, while others have taken months to disclose the incident.

The Findings

Institution Type

The PRC uses seven categories for institution type: Businesses - Other (BSO), Businesses - Financial and Insurance Services (BSF), Businesses - Retail/Merchant (BSR), Educational Institutions (EDU), Government and Military (GOV), Healthcare - Medical Providers (MED), and Nonprofit Organizations (NGO). We first analyze the number of records followed by the number of incidents in these categories. Table 1 shows the number and percentage of records breached through exposure. The government has the greatest number of records, 8,812,761, or 67.74%. Business-Other is a distant second with 10.09% followed closely by Business-Financial, 7.82%, and Medical, 6.16%. The government had two incidents that exposed over 2 million records, one at 2 million and the other at 3.4 million. The number of records exposed for a single incident ranged from 8 records to 3.4 million records. A single factor analysis of variance (ANOVA) showed no significance in the number of records per institution type ($p = 0.19416$, $\alpha=.05$).

Table 1 No. Records per Institution		
GOV	8,812,761	67.74%
BSO	1,313,052	10.09%
BSF	1,017,020	7.82%
EDU	983,173	7.56%
MED	801,433	6.16%
BSR	81,551	0.63%
NGO	1,350	0.01%

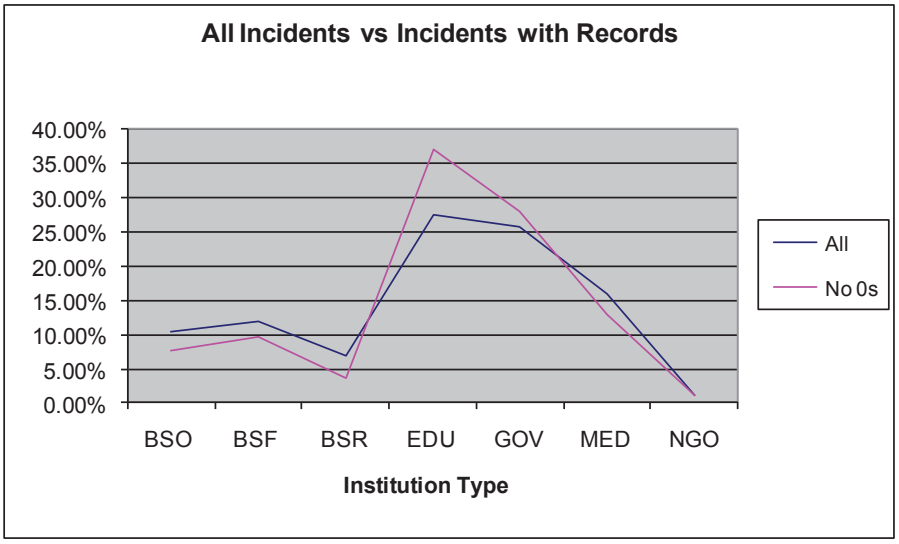
Table 2 presents the number and percentage of incidents per institution. A review of the data by institution shows that the number of incidents is not directly related to the number of records exposed. Education had the highest percentage of incidents, 37.02, but one of the lowest percentage of records exposed, 7.56. The government had the second highest number of incidents at 27.9%. A P-value of 7.88E-17 (α=.05) indicated a significant difference in the number of incidents per institution. Education and Government had a much higher number of incidents than the other institution types.

Table 2 Incidents with Record Count		
Type	**Incidents**	**Percent**
EDU	134	37.02%
GOV	101	27.90%
MED	47	12.98%
BSF	35	9.67%
BSO	28	7.73%
BSR	13	3.59%
NGO	4	1.10%

We then looked at institution type when all exposure incidents were included. That is, even those with an unknown, indeterminate or zero number of identified records. The relative percentages remain the same when comparing all incidents and only those incidents with an identified number of recorded. These results are shown in Table 3 and Figure 1. Again, Education had the highest percentage followed closely by Government.

Table 3 All Exposure Incidents		
Type	**Incidents**	**Percent**
EDU	163	27.58%
GOV	153	25.89%
MED	94	15.91%
BSF	71	12.01%
BSO	62	10.49%
BSR	41	6.94%
NGO	7	1.18%

Figure 1 All Exposure Incidents vs Incidents with Identified Number of Records



Breaches by Year

Table 4 presents the number of records and incidents by year. A longitudinal examination of the data reveals that organizations are not necessarily getting better at eliminating these types of breaches. A large breach in 2005 was the impetus for the tracking of data breaches by multiple organizations and for many states to start requiring that companies notify their customers of a data breach. The significant increase from 2005 to 2006 is therefore understandable. However, the number of incidents remained about the same from 2006 to 2008. A significant decrease in 2009 is encouraging until we see that the number of incidents increases back to the 2006-2008 level in 2010. Table 4 also highlights when the breaches with a large number of records were reported. The government had a breach of 2 million records in 2006 and of 3.4 million records in 2008. Figures 2-4 present this data graphically.

Table 4 Breaches By Year		
	Records	**Incidents**
2005	481,008	17
2006	3,292,319	76
2007	2,291,290	78
2008	4,782,582	72
2009	362,027	45
2010	1,801,114	74

Figure 2 Number of Records and Incidents by Year

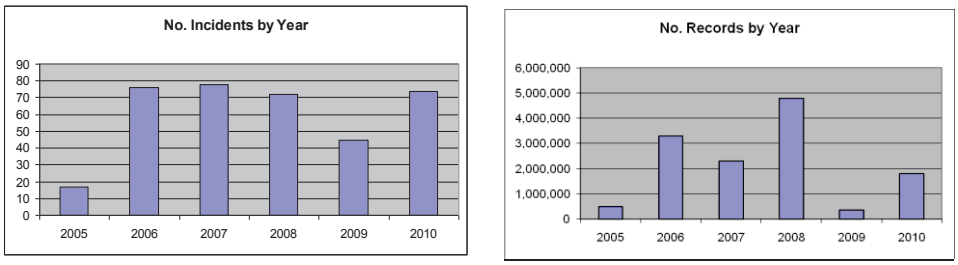


Figure 3 Percentage of Incidents by Year

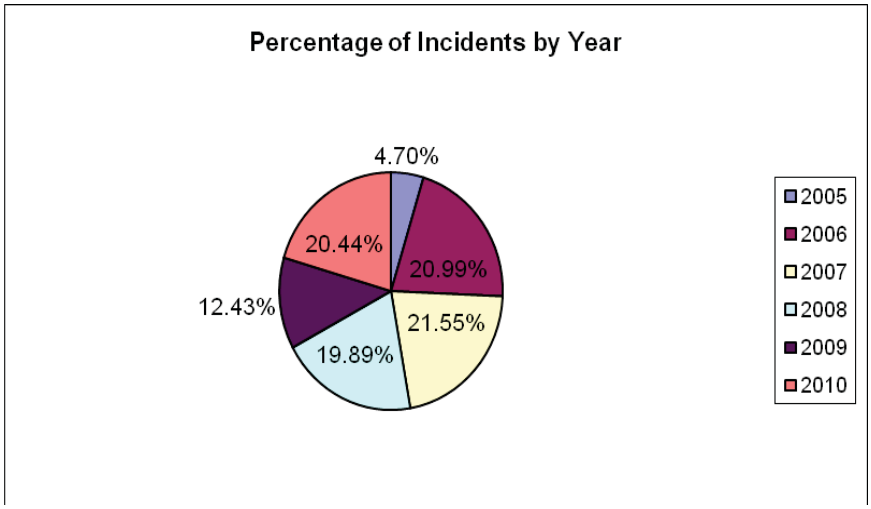
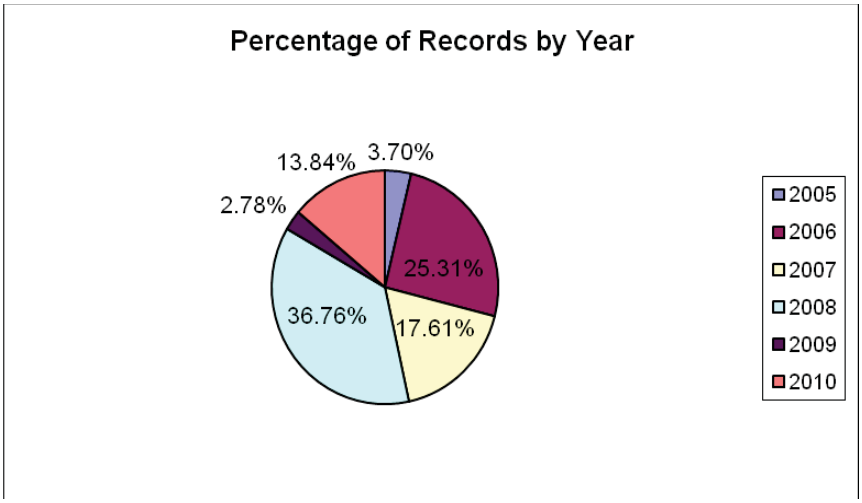


Figure 4 Percentage of Records by Year



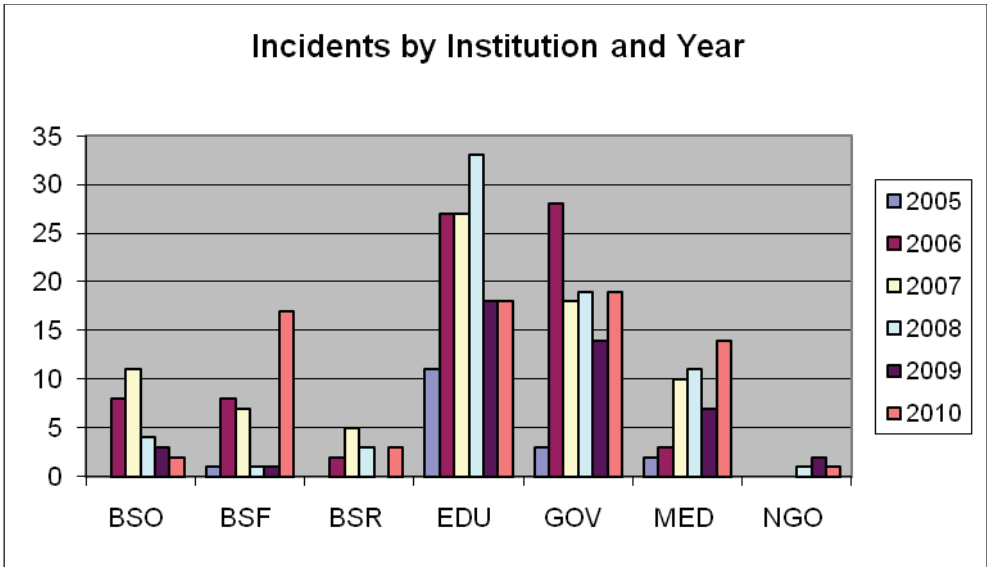
Incidents by Institution Type and Year

Table 5 presents the breached data by both institution type and year the breach was disclosed. Three institution types, BSO, BSR, and NGO, had no incidents in 2005. NGO also had no incidents in 2006 and 2007; this institution type only had 4 incidents for the six year period. The only other institution type without any reported breaches for a year since 2006 is BSR. Breaches in general occurred in every institution type in most years of the study. Figure 6 is a bar graph of incidents by institution and year. Using Table 5 and Figure 5 we can determine that Education had more incidents in 2005, 2007, 2008, and 2009 than the other institution types and had its greatest number of incidents per year, 33, in 2008. Government had more incidents in 2006 and 2010 than the other institution types and had its highest number of incidents in a year in 2006. From Figure 5, we can readily see a spike in the number of breaches by BSF, financial businesses in 2010. There is also an increase in the number of medical institutions with breaches in 2010. The number of breaches by other institution types remained about the same or decreased in 2010.

Table 5 Incidents Breached Per Year by Institution

	2005	**2006**	**2007**	**2008**	**2009**	**2010**
BSO		8	11	4	3	2
BSF	1	8	7	1	1	17
BSR		2	5	3		3
EDU	11	27	27	33	18	18
GOV	3	28	18	19	14	19
MED	2	3	10	11	7	14
NGO				1	2	1

Figure 5 Incidents by Institution and Year

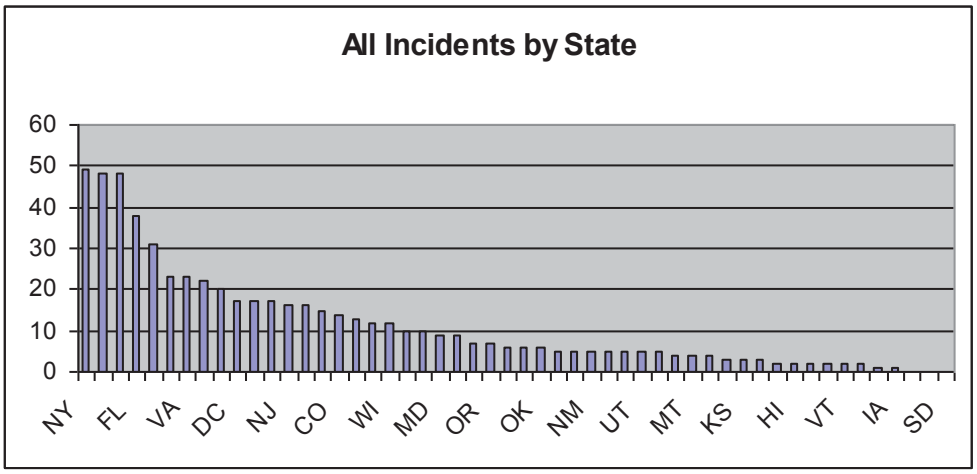


Incidents by State

The state in which the breach occurred or the headquarters of the breached organization for all recorded incidents is shown in Table 6. The data is sorted based on the number of breaches in that state. Forty-seven states and the District of Columbia had breaches through accidental exposure. New York has the greatest number of incidents at 49, followed closely by California and Texas with 48 incidents. These states also have the largest populations (US Census 2008) and the largest economies (Bureau of Economic Analysis 2010). Thirteen states had fewer than five incidents, and three states had no reported incidents, Arkansas, North Dakota and South Dakota. The data includes two incidents listed as outside the US; one with headquarters outside the US, and the other originated with a mistake by the US Consulate in Jerusalem. Figure 6 presents the pattern of incidents by state graphically

Table 6 Number Incidents (All) Per State									
ST	INC	ST	INC	ST	INC	ST	INC	ST	INC
NY	49	DC	17	MI	10	RI	5	AL	3
CA	48	PA	17	MD	9	NM	5	HI	2
TX	48	TN	16	MO	9	WA	5	ID	2
FL	38	NJ	16	OR	7	SC	5	AR	2
OH	31	GA	15	MN	7	AZ	5	VT	2
VA	23	CO	14	OK	6	ME	4	WV	2
IN	23	KY	13	MS	6	MT	4	Other	2
NC	22	WI	12	NV	6	NE	4	IA	1
IL	20	LA	12	NH	5	WY	3	DE	1
MA	17	CT	10	UT	5	KS	3	**AK, ND, SD**	

Figure 6 All Incidents by State



By Subtype

Based on an analysis of the breach descriptions, seven exposure subtypes were identified: Disposal, Email, Internet, Mail, Released, Storage, and Transport. Example Disposal incidents include records with PII being placed in dumpsters or left in unoccupied buildings. E-mail breaches include unencrypted PII of a list of customers being sent to the entire customer list. Internet includes files on public websites, on public networks, and the ability to view others information on password protected sites. Example Mail breaches include the SSN displayed on the mailing label and information sent to the wrong recipient. Released includes information provided to unauthorized recipients. Improperly stored boxes and unencrypted media are examples of Storage Breaches. Documents lost during transport because they

fell off a truck or because boxes of documents were damaged while being placed on a truck are examples of Transport. Table 7 shows the number of incidents and records breached in each subtype. Inappropriate handling of the Internet caused the greatest number of breaches, 209 or 57.73%. The large number of breached records in Storage was caused by a single incident of file boxes being left unattended for at least a month, causing the breach of 2 million records. The risk of compromised PII through exposure exists in every subtype. Figure 7 shows the incidents by subtype and institution. As might be expected by the high number of Internet incidents, this type of incident is predominant in every institution. A single factor ANOVA revealed no significant difference in the number of records per subtype ($p= 0.19586$, $\alpha=.05$); however there was a significant difference in incidents per subtype ($p = 5.99E-45$, $\alpha=.05$).

Table 7 Breaches by Subtype

	Records	Incidents	% Incidents
Disposal	518,894	35	9.67
Email	104,397	41	11.33
Internet	6,999,464	209	57.73
Mail	2,506,878	53	14.64
Released	835,680	12	3.31
Storage	2,005,389	5	1.38%
Transport	39,638	7	1.93%

Figure 7 Incidents by Institution and Subtype

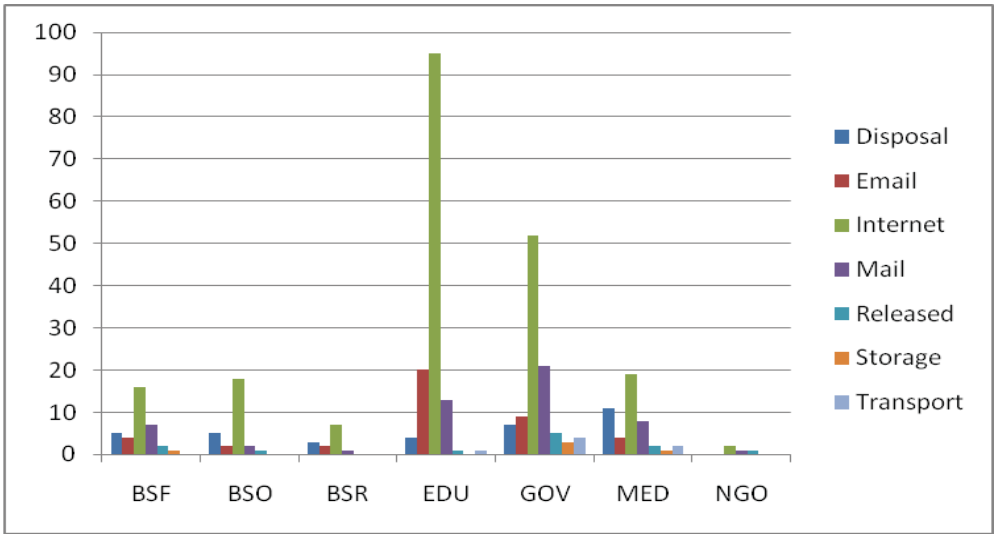


Figure 8 shows the same information as Figure 7 without the Internet incidents which dominate the scale. After the Internet, the following subtypes are most prevalent in the following institutions: BSF-mail, BSO-Disposal, BSR-Disposal, EDU-E-mail and mail, GOV-Mail, MED-Disposal, NGO-Mail and Released.

Figure 8 Incidents by Institution and Subtype, excluding Internet

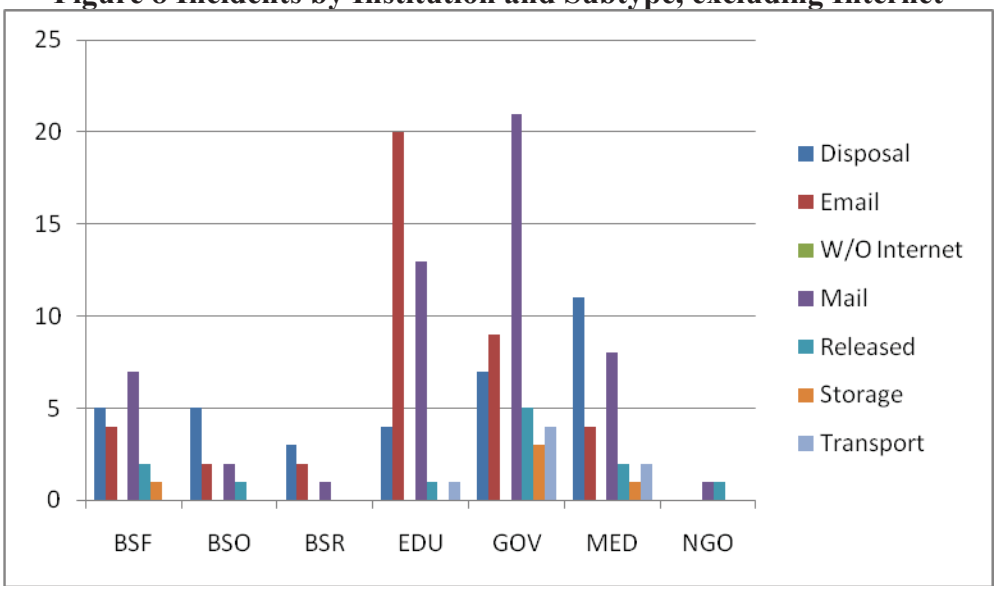


Figure 9 presents the data with institution type on the Y-axis. This figure allows one to examine which institution had more of a particular subtype. The data is presented in Figure 10, excluding the Internet incidents, to unmask the other data types. Government and Education had the greatest number of incidents so it is not unexpected that they have the greatest number of incidents in a particular category. Government had the greatest number of Mail, Released, Storage and Transport incidents. Education had the greatest number of E-mail and Internet incidents. Notice, however, that the majority of Disposal incidents were Medical. A two-factor ANOVA without replication revealed that we would accept the null hypothesis that all institutions are the same ($p = 0.058202$ $\alpha=.05$) relative to subtype.

Figure 9 Incidents by Subtype and Institution

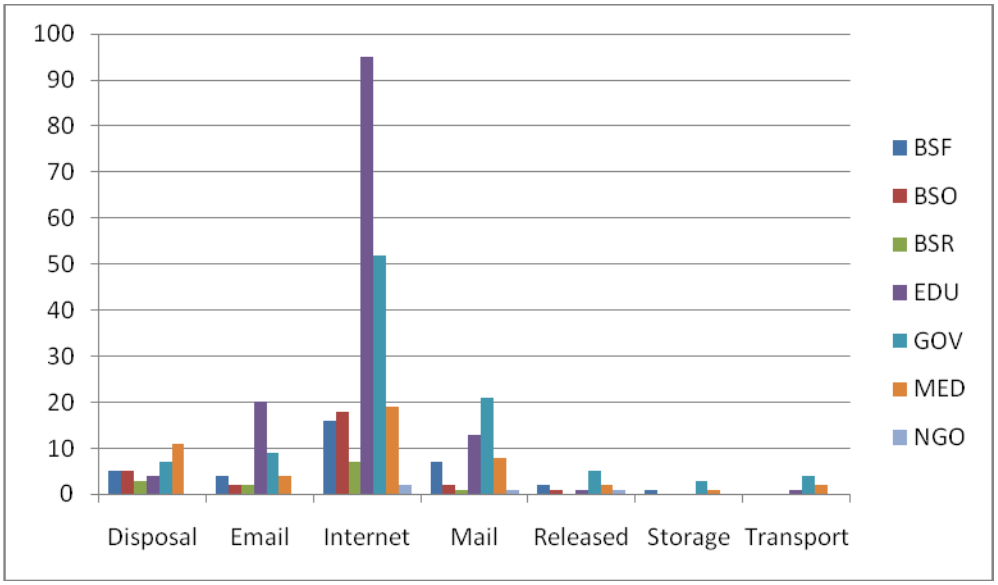
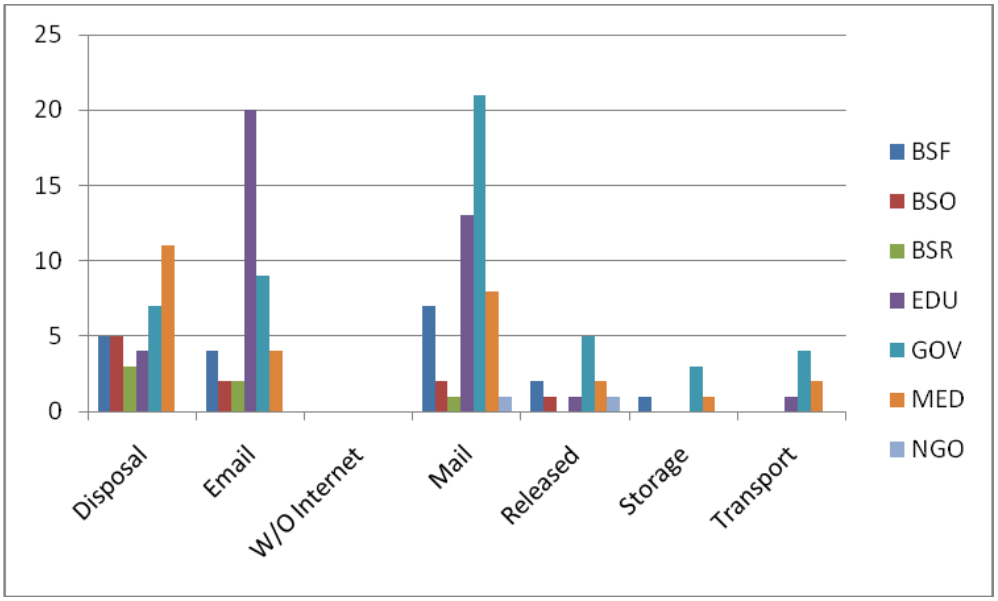


Figure 10 Incidents by Subtype (excluding Internet) and Institution



Records

Because records affect individuals, we determined the number of incidents with various record sizes. Table 8 presents these calculations. The greatest number of incidents breached between 1,000 and 10,000 records, 33.15% followed closely by the number of incidents with 100-1,000 records breached. The number of incidents with fewer than 100 records is close to the number of incidents with 10,000 to 100,000 records, 49 and 55 respectively. A regression analysis on the data resulted in an insignificant slop coefficient, suggesting an insignificant relationship between incidents and group size ($p = 0.195219817$, $\alpha=.05$).

Table 8. Incidents by Record Size

Record Size	# Incidents	Percent
<100	49	13.54
[100-1,000)	116	32.04
[1,000-10,000)	120	33.15
[10,000-100,000)	55	15.19
[100,000-1,000,000)	20	5.25
>=1,000,000	2	.55

Repeats

Twenty-four organizations had from 2-5 repeat incidents. A national organization of independent and locally operated companies that had 5 incidents in four states is not included in the table below. These five incidents involved 219,269 records. Two of these five incidents involve the same state organization and are therefore included with the repeats. Two universities and their medical facilities are also not listed because they are recorded as two different institution types, education and medical. Likewise, a state government that involved two departments, one listed as government the other as medical is not included. Table 9 lists the repeats using an identifying tag that associates the organization type with its state. If the state has multiple organizations with repeats it is designated by a sequence number, for example FL-1 and FL-2. A further analysis of the repeats revealed that fifteen organizations had repeats of the same subtype: Mail-2, E_mail-1, and Internet-12. In addition, thirteen organizations had repeats in the same year, 8 had incidents in consecutive years and 7 organizations had 3 or more incidents.

Table 9 Organizations with two or more incidents									
		2005	**2006**	**2007**	**2008**	**2009**	**2010**	**Totals**	**Records**
CA-1	GOV		2	1			1	**4**	53,154
DC-1	GOV		1	1				**2**	388,700
FL-1	BSF			2				**2**	15,266
FL-2	EDU			1	1	2	2	**6**	14,127
GA-1	GOV		2					**2**	42,000
IN-1	GOV			1		1		**2**	12,775
IN-2	EDU			3		1		**4**	1,298
KY-1	EDU		3					**3**	2,010
MT-1	EDU		1	2				**3**	314
NC-1	BSF		1	1				**2**	3,569
NY-1	BSF			1			1	**2**	605,208
NY-2	EDU	1	1					**2**	867
NY-3	EDU		1		1			**2**	5,098
NY-4	EDU		1		1			**2**	30,051
OH-2	EDU				2	1		**3**	18,542
PA-1	BSF						2	**2**	53,680
PA-2	GOV				2			**2**	32,845
PA-3	MED			2				**2**	6,088
SC-1	EDU		1	1				**2**	2,882
TX-1	EDU				2			**2**	4,430
TX-2	EDU				2		1	**3**	17,513
VA-1	GOV		2					**2**	200,000
VA-2	EDU		2					**2**	2,661
WY-1	GOV						2	**2**	14,000
Totals		**1**	**18**	**16**	**11**	**5**	**9**	**60**	**1,527,078**

Table 10 groups the repeat incidents by institution type. EDU had the greatest number of organizations with multiple incidents, 12. GOV had the second largest number of organizations with multiple incidents, 7.

Table 10 Repeats by Institution Type			
Institution	Organizations	Records	Incidents
BSF	4	677,723	8
EDU	12	99,793	34
GOV	7	743,474	16
MED	1	6,088	2

Conclusion

This paper uses the Privacy Rights Clearinghouse breach data to analyze a particular type of breach, exposure. These breaches are caused by accidents and mishandling, and should be easily preventable because they require little technical expertise to avert. An analysis of exposed data provides the following information

1. Government organizations placed the largest number of records of PII at risk through exposure; nearly 9 million records, or 67.74% of all exposure records. Business-Other was a distant second at 10.09%. The government was the only organization to expose over 1 million records with a single incident. Two incidents of 2 and 3.4 million records contributed to over half the government records exposed.
2. Education had the greatest number of incidents, 134 or 37.02%. The government was a relatively close second with 101 incidents or 27.9%. Though Education had the largest number of incidents, these organizations had the fourth largest number of records.
3. The number of incidents of this particular breach type is not declining. In four of six years, the number of incidents remained about the same, 76, 78, 72, and 74.
4. With a single exception, breaches occurred in every institution type each year since 2006. Business-Retail did not have an exposure breach in 2009. There was a spike in Business-Financial incidents in 2010, and an increase in medical incidents in 2010. Only Business-Other had a decrease in the number of incidents in 2010.
5. Breach incidents are concentrated in states with the largest populations and economies, New York, California, Texas, Florida and Ohio.
6. The overwhelming method of accidentally making PII available to the public is through the Internet. This includes being available on websites and on public servers. Nearly sixty percent, 57.73% or 209, of the exposure incidents are Internet. The next largest number is a

distant 53 or 14.64% exposed by Mail. The 209 Internet incidents compromised nearly 7 million records; 2.5 million were exposed by Mail. These results are in line with statistics maintained by Open Security Foundation's DataLossDB (2011). DataLossDB has statistics for the following four categories for all data breaches recorded: Disposal-document, Mail, E-mail and Web. Using the percentages provided and considering only those four categories, Web was the largest at 44.44%, followed by Disposal 22.22%, Mail 18.52% and Email-14.81%.

7. Internet incidents are prevalent across institution type. Disposal incidents are most prevalent by Medical, Email by Education, and Mail by Government organizations.
8. The majority of incidents exposed between 100 and 10,000 records. Thirty-two percent of incidents were between 100-1000 records and 33% between 1 and 10,000 records.
9. Twenty-four organizations had multiple incidents. Thirteen organizations had three to six incidents. Twelve, or half those with multiple incidents, were Educational organizations.

There is a significant difference in the number of incidents per institution and incidents per subtype. Government and Education should especially focus on decreasing the number of exposure breaches. All institutions should be aware that more breaches are exposed through the Internet than any other subtype.

While defending against hacking requires constant vigilance and technical expertise, lack of policy, lax policies, uninformed employees, carelessness, and negligence should not be a major source of security risk. Unfortunately, that is not the case. Equally unacceptable is that, despite continued focus on information security, organizations are not sufficiently encrypting data. The majority of incidents in this study would not be included had the data been encrypted.

References

Bureau of Economic Analysis. 2010. Gross Domestic Product by State. http://www.bea.gov/regional/gsp/ (accessed August 25, 2011).

FTC, nd. The FTC's Consumer Sentinel Network. https://www.sentinel.gov/ (accessed August 25, 2011).

FTC, 2011-1. Consumer Sentinel Network Data Book for January – December 2010, March 2011, http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2010.pdf (accessed August 25, 2011).

FTC, 2011-2. Federal Trade Commission Fiscal Year 2012 Congressional Budge Justification Summary. http://www.ftc.gov/ftc/oed/fmo/budgetsummary12.pdf (accessed August 25, 2011).

ITRC, 2011. Data Breaches http://www.idtheftcenter.org/artman2/publish/lib_survey/ITRC_2008_Breach_List.shtml# (accessed Aug 25, 2011).

Javelin Strategy & Research, 2011. 2011 Identity Fraud Survey Report: Consumer Version Prevention – Detection – Resolution. https://www.javelinstrategy.com/brochure/207 (accessed Aug 25, 2011).

Lennon, Mike, 2011. Massive Breach at Epsilon Compromises Customer Lists of Major Brands. *Security Week*. April 02, 2011. http://www.securityweek.com/massive-breach-epsilon-compromises-customer-lists-major-brands (accessed Aug 25, 2011).

Morran, Chris, 2011. E-Mail Breach Hits Best Buy, TiVo, Walgreens, Chase, Kroger, Many More. *The Consumerist*. April 4, 2011. http://consumerist.com/2011/04/e-mail-breach-hits-best-buy-tivo-walgreens-chase-kroger-many-more.html (accessed Aug 25, 2011).

Ncube, Cathy and Chlotia Garrison, 2010. Lessons Learned from University Data Breaches, *Palmetto Business and Economic Review*.13:27- 37.

Open Security Foundation, 2011-1. STATS. http://datalossdb.org/statistics (accessed Aug 25, 2011).

Open Security Foundation, 2011-2. Breach Types. http://datalossdb.org/ (accessed Aug 25, 2011).

Privacy Rights Clearinghouse, 2011. Chronology of Data Breaches Security Breaches 2005-Present. http://www.privacyrights.org/data-breach#CP (accessed Aug 25, 2011).

Talbot, David, 2011. Breached Companies Say They Did All They Could. *Technology Review*. http://www.technologyreview.com/business/37700/ (accessed Aug 25, 2011).

US Census, 2008. State Rankings. http://www.census.gov/statab/ranks/rank01.htm (accessed Aug 25, 2011).